

Temperature-Dependent Folding Pathways of Pin1 WW Domain: An All-Atom Molecular Dynamics Simulation of a G \bar{o} Model

Zhonglin Luo,* Jiandong Ding,* and Yaoqi Zhou^{†‡}

*Key Laboratory of Molecular Engineering of Polymers of Ministry of Education, Department of Macromolecular Science, Advanced Materials Laboratory, Fudan University, Shanghai, China; [†]Howard Hughes Medical Institute Center for Single Molecule Biophysics, Department of Physiology and Biophysics, State University of New York at Buffalo, Buffalo, New York; and [‡]Indiana University School of Informatics and Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, Indiana

ABSTRACT We study the folding thermodynamics and kinetics of the Pin1 WW domain, a three-stranded β -sheet protein, by using all-atom (except nonpolar hydrogens) discontinuous molecular dynamics simulations at various temperatures with a G \bar{o} model. The protein exhibits a two-state folding kinetics near the folding transition temperature. A good agreement between our simulations and the experimental measurements by the Gruebele group has been found, and the simulation sheds new insights into the structure of transition state, which is hard to be straightforwardly captured in experiments. The simulation also reveals that the folding pathways at approximately the transition temperature and at low temperatures are much different, and an intermediate state at a low temperature is predicted. The transition state of this small β -protein at its folding transition temperature has a well-established hairpin 1 made of $\beta 1$ and $\beta 2$ strands while its low-temperature kinetic intermediate has a formed hairpin 2 composed of $\beta 2$ and $\beta 3$ strands. Theoretical results are compared with other simulation results as well as available experimental data. This study confirms that specific side-chain packing in an all-atom G \bar{o} model can yield a reasonable prediction of specific folding kinetics for a given protein. Different folding behaviors at different temperatures are interpreted in terms of the interplay of entropy and enthalpy in folding process.

INTRODUCTION

Slow folding rates of β -proteins make it computationally challenging to study their folding kinetics. As a result, most simulation studies on folding of β -proteins are limited to small β -sheet domains (1–6) with a few exceptions (7–10). WW domains constitute a family of three-stranded antiparallel β -sheet proteins comprised of ~ 34 –44 amino acids (11–14). They are named after two highly conserved tryptophans in the strands. Some WW domains are suitable for theoretical folding studies (15–22) because they unfold and refold reversibly and cooperatively (two-state kinetics) (23,24). Moreover, they are amenable to well-defined mutagenesis without significant changes in their structures (16,25–27). Protein Pin1 is a two-domain protein that regulates early mitotic events. Its WW domain binds proline-rich peptides, and another domain of Pin1 catalyzes the proline *cis/trans*-isomerization. The high-resolution x-ray structure of the Pin1 WW domain has been determined (1.35 Å) and only residues 6–39 are visible (28). The domain contains two hydrophobic clusters at the opposite side of strands. Core 1 contains Leu⁷, Trp¹¹, Tyr²⁴, and Pro³⁷; Core 2 has Tyr²³, Phe²⁵, and the *n*-propyl portion of Arg¹⁴, as shown in Fig. 1 A. The NMR structure of this domain adopts a more flexible loop 1 (residues Ser¹⁶ to Arg²¹) in solution compared to the x-ray structure (29).

The folding transition of Pin1 WW domain has been studied intensively by experimental techniques. A series of

well-designed side-chain mutagenesis by Gruebele group (26) suggested that the locally interacting loop 1 controls the folding rate of the Pin1 WW domain with nonclassical ϕ -values > 1 at the physiological temperature, while the formation of the hydrophobic Core 1 contributes significantly to thermal stability. The main-chain amide-to-ester (A-to-E) mutation studies of the Pin1 WW domain also demonstrated that its folding transition state (TS) has natively like main-chain hydrogen bonds (H-bonds) in loop 1, less developed H-bonds in loop 2 and other regions at 50°C, which is close to the midpoint temperature of thermal denaturation ($T_m = 59^\circ\text{C}$) (27).

Pin1 WW domain has also been studied theoretically. Karanicolas et al. compared the folding kinetics and thermodynamics of three WW domains and found that the difference in hydrophobic packing leads to different folding behaviors (biphasic kinetics of FBP WW domain versus two-state kinetics of Pin1 WW domain and hYAP WW domain) (17,18). Brown et al. (30) refolded proteins L and G with a three-letter code sequence (the hydrophobic, hydrophilic, and neutral) and found that this model could also lead to a correct folding of the Pin1 WW domain. Different folding pathways, however, have been suggested by different theoretical studies. Karanicolas et al. proposed that the hairpin $\beta 2$ - $\beta 3$ (hairpin 2) of the Pin1 WW domain folds before hairpin $\beta 1$ - $\beta 2$ (hairpin 1) and the hydrophobic Core 1 forms at last (19) by an improved C_α model including nonnative interactions. More recently, Cecconi et al. studied the transition state ensemble (TSE) of the Pin1 WW domain by a C_α -based model with three-letter codes and found that the folding of the Pin1 WW domain was initiated from the

Submitted November 28, 2006, and accepted for publication May 10, 2007.

Address reprint requests to J. Ding, E-mail: jdding1@fudan.edu.cn; or Y. Zhou, E-mail: yqzhou@iupui.edu.

Editor: Jose Onuchic.

© 2007 by the Biophysical Society

0006-3495/07/09/2152/10 \$2.00

doi: 10.1529/biophysj.106.102095

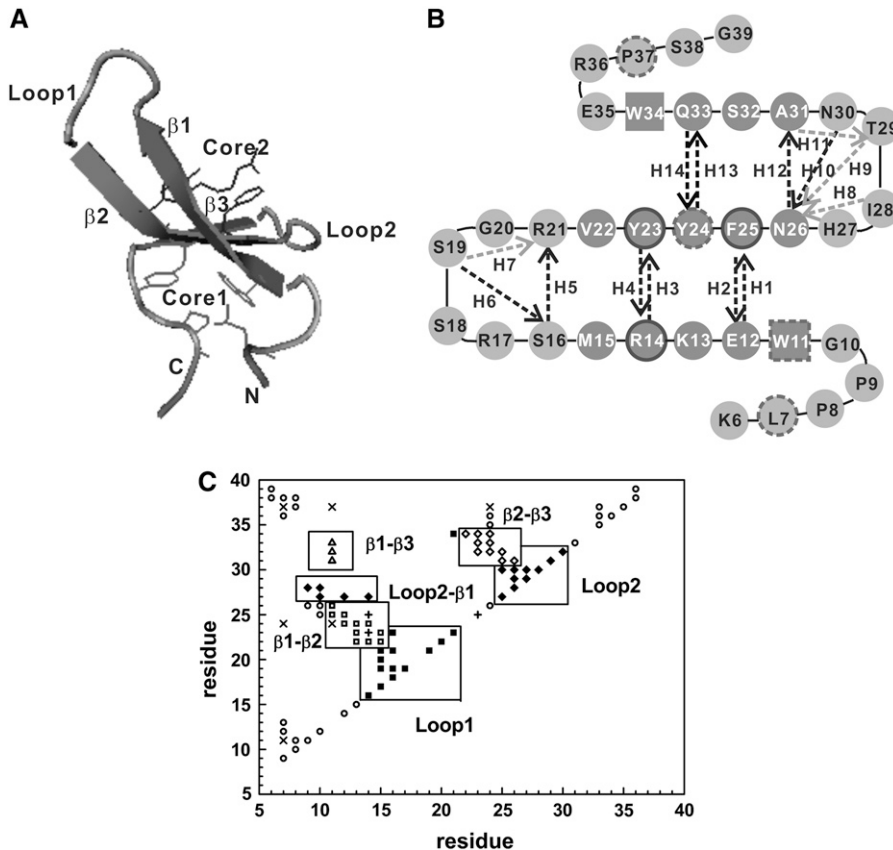


FIGURE 1 (A) A cartoon display of the global-minimum-energy structure of an all-atom off-lattice model of Pin1 WW domain (Lys⁶ to Gly³⁹). Three β -strands, two loop regions and coil regions, the hydrophobic Core 1 (Leu⁷, Trp¹¹, Tyr²⁴, Pro³⁷), and Core 2 (Arg¹⁴, Tyr²³, Phe²⁵) are specifically indicated. Drawn with VMD (65). (B) Two-dimensional presentation of the Pin1 WW domain. In addition to the secondary-structure elements, the two conserved Trp residues are highlighted by squares. Ten main-chain H-bonds and four side-chain H-bonds are also shown by the dark dashed lines and the light dashed lines, respectively, and each H-bond is indicated by an arrow from a hydrogen donor to the associated hydrogen acceptor. Core 1 is highlighted with a dashed-line border, while Core 2, with a solid-line border. (C) The contact map of the native structure. A residue is in contact with another residue if there is at least one square-well atomic contact between them. Square symbols denote $\beta 1$ - $\beta 2$ contacts; diamonds, $\beta 2$ - $\beta 3$ contacts; triangles, $\beta 1$ - $\beta 3$ contacts; crosses, contacts among Core 1 residues; pluses, contacts among Core 2 residues; the solid squares indicate contacts between loop 1 residues and other residues; the solid diamonds refer to contacts between loop 2 residues and other ones; and other contacts are labeled with circles. The total number of native residue-residue contacts ($|i-j| \geq 2$) is 91. The number of nonlocal native contacts ($|i-j| > 2$) is 65.

hydrophobic contacts in $\beta 1$ - $\beta 2$ and $\beta 2$ - $\beta 3$ and completed by zipping the hairpin $\beta 1$ - $\beta 2$ and hairpin $\beta 2$ - $\beta 3$ toward loop 1 and loop 2, respectively (20).

In this article, we make an in-depth study of the folding pathway of the Pin1 WW domain based on an all-atom $G\ddot{o}$ model (31). The constant-temperature discontinuous molecular dynamics (MD) simulation was employed to examine the thermodynamic and kinetic behaviors of the Pin1 WW domain. The simulation technique and the associated $G\ddot{o}$ model have been used successfully to simulate folding thermodynamics and kinetics of a three-helix bundle and a β -hairpin (32–35). Here, we use the discontinuous MD technique because a discontinuous MD simulation with a square-well potential is computationally more efficient than a normal MD simulation for the following reasons. First, only short-range interactions are involved; second, force calculations are not required in the discontinuous MD simulations; and third, particles move at constant velocities between collisions and there is an efficient search algorithm for the collision time.

Here, we found that there are two distinct folding kinetic pathways at different temperatures for the Pin1 WW domain. At the transition temperature, Pin1 WW domain folds without any intermediates, and its transition state is characterized by a mostly formed hairpin 1. At a lower temperature, however, Pin1 WW domain folds via a kinetic intermediate

with a mostly established hairpin 2. Theoretical results are compared with available experimental data.

METHODS AND SIMULATIONS

Model

The detail description of the discontinuous MD method could be found elsewhere (32–35). Here, we give a brief summary of the model in simulation. First, all the heavy-atom positions of the WW domain were acquired from the x-ray structure (residues 6–39 of Protein DataBank 1pin) (28). The N-terminus and *n*-methyl C-terminus of the polypeptide were capped with acetyl and amine groups, respectively. The CHARMM program (36) was used to create the initial positions of polar hydrogen atoms. Steric clashes were removed by 100 minimization steps with the CHARMM polar hydrogen parameter set 19 and distance-dependent dielectric constants (37). Thus, all heavy atoms and polar hydrogens are included in the model and the total number of atoms is 361.

We constrain two bonded atoms, as well as any 1,3 angle-constrained atomic pair and 1,4 aromatic carbon pair, i and j , to a center-to-center distance between $0.9\sigma_{ij}$ and $1.1\sigma_{ij}$, where σ_{ij} is the atomic distance obtained by the CHARMM energy minimization as mentioned above. This constraint is denoted by an infinitely deep square-well potential

$$u_{ij}^{\text{bond}}(r) = \begin{cases} \infty, & r < 0.9\sigma_{ij} \\ 0, & 0.9\sigma_{ij} < r < 1.1\sigma_{ij} \\ \infty, & r > 1.1\sigma_{ij} \end{cases} \quad (1)$$

We apply a bondlike potential for improper dihedral angels to maintain the L-form chirality about a tetrahedral extended heavy atom

($\omega_0 = 35.26439^\circ$) and to keep certain planar atoms ($\omega_0 = 0^\circ$). A 20° fluctuation is used to increase the flexibility and folding rate. That is,

$$u_{\omega}^{\text{improp}} = \begin{cases} \infty, & \omega - \omega_0 > 20^\circ \\ 0, & -20^\circ < \omega - \omega_0 < 20^\circ \\ \infty, & \omega - \omega_0 < -20^\circ \end{cases} \quad (2)$$

The nonbonded i, j pair interactions are represented by a hard-core and square-well potential

$$u_{ij}(r) = \begin{cases} \infty, & r < 0.8\sigma_{ij}^{\text{vdW}} \\ B_{ij}\epsilon, & 0.8\sigma_{ij}^{\text{vdW}} < r < 1.2\sigma_{ij}^{\text{vdW}} \\ 0, & r > 1.2\sigma_{ij}^{\text{vdW}} \end{cases} \quad (3)$$

where σ_{ij}^{vdW} are the van der Waals parameters from the CHARMM polar hydrogen parameter set 19 (37) and B_{ij} are the nonbonded bead interaction strengths. The factor of 0.8 is the typical ratio between the hard-core diameters and the van der Waals parameters in the Weeks-Chandler-Andersen perturbation theory (38), while a ratio of 1.5 (= 1.2:0.8) between the square-well and hard-core diameters is typical for small molecular systems (39).

We remove all initial hard-core overlaps in the CHARMM minimized structure by a short discontinuous MD simulation. To ensure that the original native contacts found in the x-ray structure are maintained during the process, the square-well interactions for these native contacts are set to be a large negative energy (-100). The initial hard-core diameter between any two overlapped atoms is regarded as the distance found in the energy-minimized structure. We modify the hard-core diameters at each time step until the true hard-core diameters are reached and the simulation continues until all initial overlaps are removed. The resulting structure has a main-chain root-mean-squared deviation (RMSD) from the x-ray structure of only 0.69 Å and is regarded as the global minimum structure for the protein (Fig. 1 A). In all, there are 2398 square-well atom-atom overlaps (including both backbone and side-chain contacts) in the global minimum structure. Henceforth, these shall be regarded as native atomic contacts.

To make sure that the energy of the native structure (shown in Fig. 1 A) is at the global minimum, a Gō interaction (31) is adopted. In a Gō model, a square-well interaction energy between two atoms is $-\epsilon$ ($B_{ij} = -1$) for a native contact, and 0 ($B_{ij} = 0$), otherwise. Henceforth, the internal energy, E^* ($= E/\epsilon$), and the reduced temperature, T^* ($= k_B T/\epsilon$, in which k_B is Boltzmann constant), are scaled in units of ϵ . A reduced time unit t^* is also used [$t^* = t(\epsilon/M\sigma_L^2)^{1/2}$ where M is the average mass of the atoms and σ_L is 1 Å]. The total number of native residue-residue contacts is 91 (the neighbor contacts between i and $i \pm 1$ are omitted, shown in Fig. 1 C), and the number of nonlocal native contacts ($|i-j| > 2$) is 65.

The formation of an H-bond is defined by a cutoff distance of 2.88 Å between a donor H atom and an acceptor O atom. Ten backbone H-bonds of the native structure are labeled in Fig. 1 B, which are the same as those shown in Fig. 1 of Jäger et al. (26) and in Fig. 1 of Deechongkit et al. (27); other native side-chain H-bonds are mainly located in two loop regions (26).

Simulations

The discontinuous MD simulations of a single Pin1 WW domain were carried out in the canonical ensemble. Ghost-solvent particles were employed to maintain a constant temperature by their collisions with protein atoms (40–42). Simulations were conducted at 18 temperatures from $T^* = 2.5$ to $T^* = 4.8$ for the thermodynamic analysis. The thermodynamic properties at the temperatures that are not simulated can be obtained through the weighted histogram method (43), which is a least-square optimization method for extracting the degeneracy factors of the energy levels and the partition function Z from multiple simulations at different temperatures. The energy and the heat capacity can be calculated from the partition function Z as

$$E^* = (T^*)^2 \frac{\partial \ln Z}{\partial T^*}, \quad (4)$$

$$C_v^* = \partial E^* / \partial T^*. \quad (5)$$

We investigated the folding kinetics at $T^* = 2.7, 3.0$, and 3.6. For folding kinetics at $T^* = 2.7$ and 3.0, the initial coil-like configurations and velocities were obtained from equilibrium simulations at $T^* = 4.0$. Seventy-seven and 135 kinetic simulations were performed for $t^* = 200,000$ at $T^* = 2.7$ and 3.0, respectively. The coordinates were recorded every 100 reduced time units for later analysis. For folding kinetics at $T^* = 3.6$, both folding and unfolding behaviors were observed at equilibrium simulation. The coordinates were recorded every 50 reduced time units. A progress variable Q , the fraction of nonlocal native contacts ($|i-j| > 2$) between residues (total number is 65, shown in Fig. 1 C), is used to monitor the folding process. Two residues are regarded as in contact if there are any square-well atomic overlaps between them.

ϕ -value

The experimental data for analyzing the folding transition state are ϕ -values obtained from protein engineering experiments (44). These experiments measured the change in the active energy of the transition state relative to the change in the free energy of the native state upon a perturbative mutation (substitution). The ϕ -values are often interpreted as the fraction of native structure or the so-called nativeness of each residue in the folding transition state. Thus, one way to predict ϕ -values is to calculate the fraction of native contacts in the TSE (45–47). That is,

$$\phi_i^{\text{sim}} = \frac{\langle N_i \rangle^{\text{TS}}}{\langle N_i \rangle^{\text{native}}} = \frac{\left\langle \sum_{j \neq i, i \pm 1} N_{ij} \right\rangle^{\text{TS}}}{\left\langle \sum_{j \neq i, i \pm 1} N_{ij} \right\rangle^{\text{native}}}, \quad (6)$$

where $\langle N_i \rangle^{\text{native}}$ and $\langle N_i \rangle^{\text{TS}}$ are the average of native heavy atomic contacts between residue i and other residues in the native condition and that at the transition state, respectively. A contact is considered to exist when the distance of any heavy atoms of the two nonadjacent residues is less than the square-well width (pairs $i, i \pm 1$ are excluded). Our Gō model does not have attractive interactions for nonnative interactions; however, it does have atomic packing interactions from nonnative contacts. Considering the possibility that the nonnative packing might also contribute to ϕ -values, we also calculate

$$\phi_i^{\text{tc}} = \frac{\langle N_i^{\text{total}} \rangle^{\text{TS}}}{\langle N_i^{\text{total}} \rangle^{\text{native}}} = \frac{\left\langle \sum_{j \neq i, i \pm 1} N_{ij}^{\text{total}} \right\rangle^{\text{TS}}}{\left\langle \sum_{j \neq i, i \pm 1} N_{ij}^{\text{total}} \right\rangle^{\text{native}}}, \quad (7)$$

where $\langle N_i^{\text{total}} \rangle^{\text{native}}$ and $\langle N_i^{\text{total}} \rangle^{\text{TS}}$ are the average of all heavy atomic contacts between residue i and other residues in the native condition and that at the transition state, respectively. In this article, $\langle N_i^{\text{total}} \rangle^{\text{native}}$ and $\langle N_i \rangle^{\text{native}}$ were calculated from the configurations of equilibrium simulations at $T^* = 2.7$.

RESULTS

Thermodynamics

The heat capacity as a function of temperature for Pin1 WW domain is shown in Fig. 2 A. A sharp peak is observed at $T^* = 3.60$, which indicates a thermodynamic transition. This transition is, according to the weighted histogram analysis, accompanied by a bimodal distribution of energies (Fig. 2 B). Thus, the observed transition is a first-order-like two-state

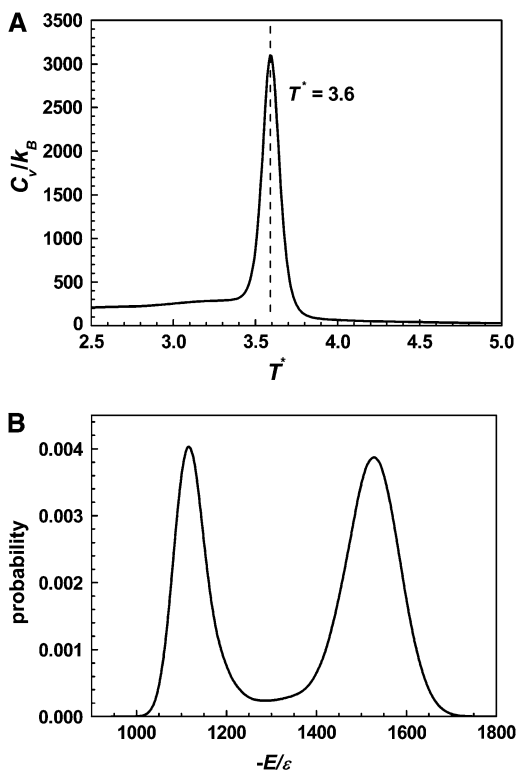


FIGURE 2 (A) Reduced heat capacity ($C_v^* = C_v/k_B$) as a function of reduced temperature ($T^* = k_B T/\epsilon$). The folding transition temperature is near $T^* = 3.60$. (B) Probability distribution of energy at the transition temperature $T^* = 3.59$. The results were obtained from the weighted histogram method.

transition. Analysis of the structures further demonstrates that the underlying transition is a folding transition between coil-like structures (the average RMSD of main-chain atoms at $T^* \geq 4.0$ is $>14 \text{ \AA}$) and folded structures (the average RMSD of main-chain atoms at $T^* \leq 3.4$ is $<3.0 \text{ \AA}$).

The two-state behavior observed above is also reflected from the free energy profiles as a function of reduced energy at different temperatures (Fig. 3). At the transition temperature ($T^* = 3.6$), there are two stable states corresponding to an unfolded state (with a reduced energy $E^* \sim -1110$) and a folded state ($E^* \sim -1530$), respectively. The transition state has an energy E^* at $\sim -1290 \pm 30$. When temperature increases, the unfolded state becomes more stable than the native state as expected. Fig. 3, A and C, also illustrate that only one thermodynamically stable state is observed at a very high temperature or a very low temperature.

If the reduced energy is a correct order parameter, a transition state can be defined based on the peak between the coil and folded states. Fig. 3 further indicates that the transition state becomes more nativelike as temperature increases (see, for instance, from $E^* \sim -1250$ at $T^* = 3.45$ to $E^* \sim -1350$ at $T^* = 3.7$), exhibiting the Hammond behavior (48), while the folded state becomes less stable (E^* from -1570 at $T^* = 3.45$ to -1500 at $T^* = 3.7$), which is consistent with the

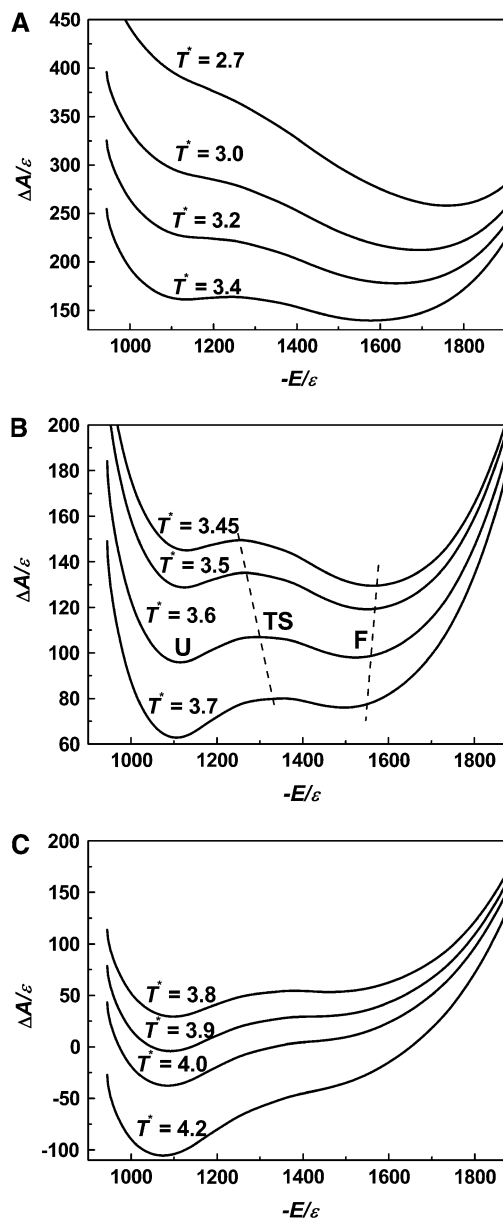


FIGURE 3 Free energy profiles as a function of reduced energy (A) at low temperatures; (B) at approximately the folding transition temperature; (C) at high temperatures. The positions of the transition states and native states change with temperature, as indicated by the dashed lines. U, unfolded state; TS, transition state; F, folded state.

ground state effects for proteins (49). (Energy can be used very well to determine the location of transition states at the transition temperature. However, it cannot be used for determining transition states at low temperatures because a kinetic intermediate exists; see below.)

Folding kinetics at the transition temperature

Fig. 4 plots the probability distribution of the number of native contacts between $\beta 1$ and $\beta 2$ residues and that between

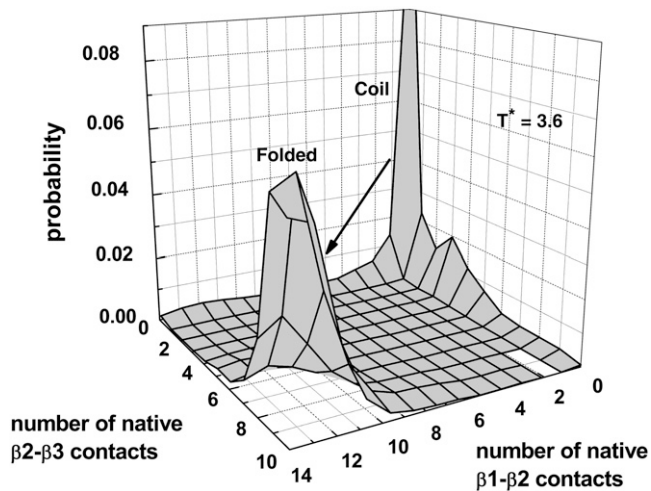


FIGURE 4 Probability distribution of the number of native contacts between $\beta 1$ and $\beta 2$ residues and that between $\beta 2$ and $\beta 3$ residues. A total of 5,130,000 configurations of eight trajectories at $T^* = 3.6$ were collected for statistics.

$\beta 2$ and $\beta 3$ residues at the transition temperature $T^* = 3.6$. At this temperature, no detectable kinetic intermediate is observed (Fig. 4). This is true when the distribution is plotted against some other progress variables.

A typical folding and unfolding trajectory is shown in Fig. 5 A, where reduced energy E^* , the fraction of nonlocal native contacts Q , and RMSD in various regions are plotted as a function of time. A series of snapshots in the folding process are shown in Fig. 5 B. The folding process starts from the local contacts of loop 1 that initiates the formation of $\beta 1$ and $\beta 2$ hairpin. This is followed by three native hydrophobic contacts of Core 2 and the stable hairpin $\beta 1$ - $\beta 2$. Then, local contacts of loop 2 bring $\beta 2$ - $\beta 3$ together, to yield the final folded state.

Transition state at the transition temperature

In this article, a snapshot or a conformation is defined as a transition state if its reduced energy is within -1290 ± 30 (Figs. 2 B and 3) and it also belongs to the region of an abrupt structural change in a folding or unfolding event in Fig. 5 A (50). This method of determining TSE is equivalent to the criterion of Q (51–53). We have also tried several other methods including the use of multiple progress variables and the folding probability along a trajectory (P_{fold}) (54). The resulting ensembles are essentially the same. Details will be discussed in a separate article. Here, we present transition state structures generated by the energy criterion only.

The samples in the TSE are made of 883 configurations collected from the folding/unfolding transitions. (We use the structures in both folding/unfolding transitions because a careful analysis indicates that the structures are the same regardless of folding or unfolding events.) The 883 configurations have an average fraction of nonlocal native contacts $Q = 0.27 \pm 0.05$ and an average radius of gyration $R_g = 13.7 \pm 1.4 \text{ \AA}$. The average main-chain-heavy-atom RMSD and all-heavy-atom RMSD values are $9.33 \pm 1.65 \text{ \AA}$ and $10.24 \pm 1.62 \text{ \AA}$, respectively. The RMSDs of main-chain heavy atoms and all heavy atoms from the global minimal structure as a function of residue index number are shown in Fig. 6. Hairpin 1 and loop 1 have much smaller deviations from the global minimum structure than loop 2 and $\beta 3$. Both N- and C-terminus are considerably flexible. Some typical snapshots of TSE are shown in Fig. 7. All of them have a well-established hairpin 1 but a yet flexible loop 1, loop 2 and $\beta 3$.

Table 1 shows the probabilities of the native H-bonds that are formed in TSE. The probabilities of the four backbone H-bonds between $\beta 1$ and $\beta 2$ strands (H1 to H4) are much higher than those of the two backbone H-bonds between $\beta 2$ and $\beta 3$ strands (H13 and H14). This confirms that hairpin 2

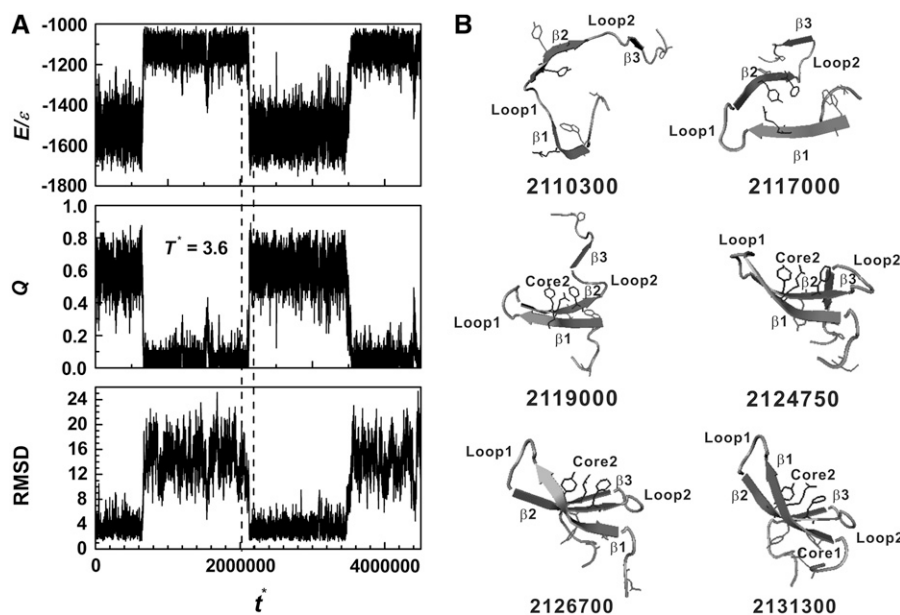


FIGURE 5 A folding trajectory at $T^* = 3.6$. (A) The reduced energy (E/ϵ), the fraction of native contacts between residues (Q), and the RMSD of Pin1 WW domain are plotted as a function of reduced time (t^*). (B) A series of snapshots at the marked reduced times which are in a folding regime indicated by the dashed lines in panel A. The snapshots are displayed with the same symbols as those in Fig. 1 A.

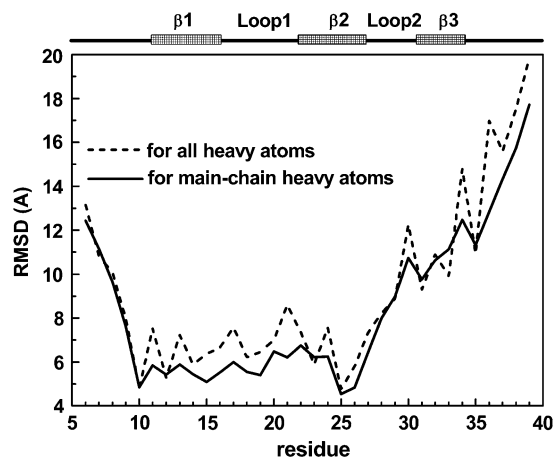


FIGURE 6 Average main-chain-heavy-atom RMSD (solid line) and all-heavy-atom RMSD (dashed line) of the transition-state structures from the native structure for each residue.

is not formed in TSE whereas hairpin 1 is well established. The H-bonds in the loop 1 region have higher values: 0.25 for H6 (H-bond between the N-H of Ser¹⁹ and the C=O of Ser¹⁶), 0.06 for H7 (H-bond between the O-H of the Ser¹⁹ side chain and the N-H of the Arg²¹ main chain); and the probability of H5 between the N-H of Ser¹⁶ and the C=O of Ser²¹, the two residues that define the loop 1, is 0.27. In contrast, the probability in loop 2 is only 0.02 for H10 (H-bond between the N-H of Asn³⁰ and the C=O of Asn²⁶); and 0.05 for H12 (H-bond between the N-H of Asn²⁶ and the C=O of Ala³¹). The probabilities of the three local side-chain H-bonds H8, H9, H11 in the loop 2 region are 0.41, 0.25, and 0.05, respectively. Thus, loop 2 has only local native contacts. The probabilities of the four backbone H-bonds between $\beta 1$ and $\beta 2$ strands are higher than the

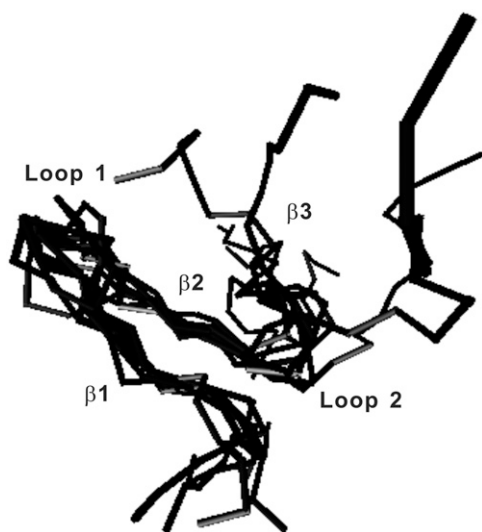


FIGURE 7 Ten snapshots in the transition-state ensemble at $T^* = 3.6$.

TABLE 1 Probability of H-bond formation in TSE at $T^* = 3.60$ (near transition temperature)

Serial*	Donor [†]	Acceptor [†]	Probability	Location
H1	Glu ¹²	Phe ²⁵	0.69	$\beta 1$ - $\beta 2$
H2	Phe ²⁵	Glu ¹²	0.51	
H3	Arg ¹⁴	Tyr ²³	0.37	
H4	Tyr ²³	Arg ¹⁴	0.46	
H5	Ser ¹⁶	Arg ²¹	0.27	Loop 1
H6	Ser ¹⁹	Ser ¹⁶	0.25	
H7	Ser ¹⁹ (O-H)	Arg ²¹ (N-H) [‡]	0.06	
H8	Ile ²⁸	Asn ²⁶ (O=CNH ₂)	0.41	Loop 2
H9	Thr ²⁹	Asn ²⁶ (O=CNH ₂)	0.25	
H10	Asn ³⁰	Asn ²⁶	0.02	
H11	Ala ³¹	Thr ²⁹ (O-H)	0.05	
H12	Asn ²⁶	Asn ³¹	0.05	
H13	Tyr ²⁴	Gln ³³	0.00	$\beta 2$ - $\beta 3$
H14	Gln ³³	Tyr ²⁴	0.00	

*Indicated in Fig. 1 B.

[†]The side-chain groups are shown in brackets.

[‡]Backbone N-H.

probabilities of H-bonds in loop 1 region, which means the H-bonds of the loop 1 are moderately stable compared to H-bonds in hairpin 1.

ϕ -value analysis

The ϕ -values obtained from our simulations are shown in both Fig. 8 and Table 2. The data were calculated according to the fraction of the number of native heavy atomic contacts (Eq. 6). The correlation coefficient between ϕ^{sim} and ϕ^{exp} (side-chain mutation) is 0.39. There is one outlier caused by the negative experimental ϕ -value for Ile²⁸. Removing this residue improves the correlation coefficient between ϕ^{sim} and ϕ^{exp} from 0.39 to 0.50. It is of interest to note that the

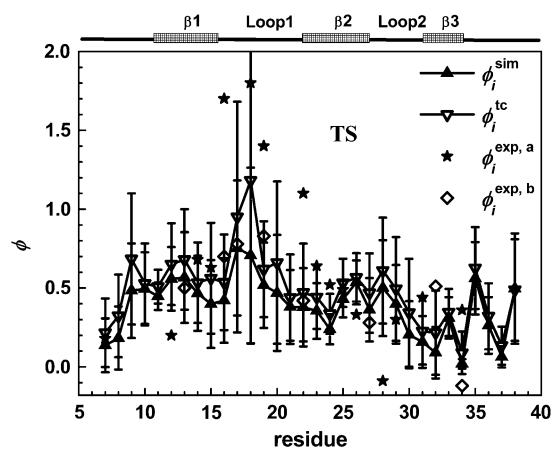


FIGURE 8 The simulated ϕ -values ϕ_i^{sim} (just native contacts involved) and ϕ_i^{tc} (containing total contacts) of each residue. The experimental $\phi_i^{\text{exp,a}}$ from side-chain mutation and $\phi_i^{\text{exp,b}}$ from main-chain mutation (A-to-E mutation) both measured by Gruebele group (26,27) are also shown. The secondary-structure elements are indicated above the diagram.

TABLE 2 Contact residues in TSE and ϕ -values from simulations (ϕ_i^{sim}) and experiments (ϕ_i^{exp})

Residue*	Contact residues [†]	ϕ_i^{sim}	$\phi_i^{\text{exp}\ddagger}$	$\phi_i^{\text{exp}\S}$	Location
L7	P9, W11	0.14	0.18		Nonsheet
P8	G10, W11	0.18			
P9	L7, W11, N26, I28	0.48			
G10	P8, E12, F25, N26, H27, I28	0.49			
W11	L7, P8, P9, Y24, F25, N26	0.45			β 1
E12	G10, R14, Y24, F25, H27	0.56	0.2		
K13	M15, V22, Y23, Y24	0.57		0.50	
R14	E12, S16, V22, Y23, Y24, F25	0.47	0.68		
M15	K13, R17, S19, G20, R21, V22, Y23	0.40	0.63		
S16	R14, S18, S19, R21, Y23	0.42	1.7	0.70	Loop 1
R17	M15, S19	0.75		0.78	
S18	S16	0.71	1.8		
S19	M15, S16, R17, R21	0.52	1.4	0.83	
G20	M15, V22	0.47			
R21	M15, S16, S19, Y23	0.38			
V22	K13, R14, M15, G20	0.38	1.1	0.42	β 2
Y23	K13, R14, M15, S16, R21, F25	0.35	0.64		
Y24	W11, E12, K13, R14, N26	0.24	0.52		
F25	G10, W11, E12, R14, Y23, H27, N30, A31, S32	0.43	0.49		
N26	P9, G10, W11, Y24, I28, T29, N30, A31	0.54	0.33		
H27	G10, E12, F25, T29, N30	0.36		0.28	Loop 2
I28	P9, G10, N26, N30	0.50	-0.09		
T29	N26, H27, A31	0.40	0.3		
N30	F25, N26, H27, I28, S32	0.20			
A31	F25, N26, T29, Q33	0.16	0.44		β 3
S32	F25, N30	0.09		0.51	
Q33	A31, E35, R36, P37	0.31			
W34	R36	0.02	0.36	-0.12	
E35	Q33, P37	0.56			Nonsheet
R36	Q33, W34, S38, G39	0.26			
P37	Q33, E35	0.06			
S38	R36	0.50			

The boldface type indicates that the associated ϕ -value(s) is (are) experimentally available.

*Name of a residue is abbreviated with a single-letter and a serial number.

[†]Native contacts in TSE with probabilities >0.1 . A residue is in contact with another one if there is at least one square-well heavy atomic contact between them.

[‡]Results from side-chain mutation (26).

[§]Results from A-to-E mutation (27).

theoretical ϕ -values have a better agreement with experimental ϕ -values based on main-chain A-to-E mutation. The correlation coefficient between ϕ^{sim} and ϕ^{exp} of A-to-E results is 0.74 for eight experimental points.

It is of interest to note that some nonnative contacts (due to nonnative side-chain packing in this article) are found in TSE. The ϕ^{tc} values based on total contacts (including both native and nonnative contacts) have a somewhat better agreement with ϕ^{exp} values. The correlation coefficient between ϕ^{tc} and ϕ^{exp} is improved to 0.53 (for experimental side-chain mutation) and 0.75 (for experimental A-to-E mutation), compared to 0.39 and 0.74 between ϕ^{sim} and ϕ^{exp} , respectively. The main difference between ϕ^{sim} and ϕ^{tc} comes from two residues (Arg¹⁷ and Ser¹⁸). It is interesting that both ϕ^{tc} and ϕ^{exp} of side-chain mutation have nonclassical ϕ -values ($\phi > 1$) at loop 1. Although our study implies that nonnative contacts in

the transition states might play a role in folding, the relation between nonclassical ϕ -values and nonnative attractive interactions is still an open question, because in a $G\bar{o}$ model, only native contacts are attractive. Further studies are needed for the effect of nonnative attractive interactions.

Folding kinetics at $T^* = 2.7$

Fig. 9 plots the probability distribution of the number of native β 1- β 2 residue contacts and the number of native β 2- β 3 residue contacts at a low temperature $T^* = 2.7$. At this temperature, we found a kinetic intermediate with a mostly formed hairpin β 2- β 3 (with 9~10 nonlocal native β 2- β 3 contacts). At a slightly higher temperature $T^* = 3.0$, a hairpin β 1- β 2 intermediate was observed in addition to the hairpin β 2- β 3 intermediate. Twenty-two out of 102 folded

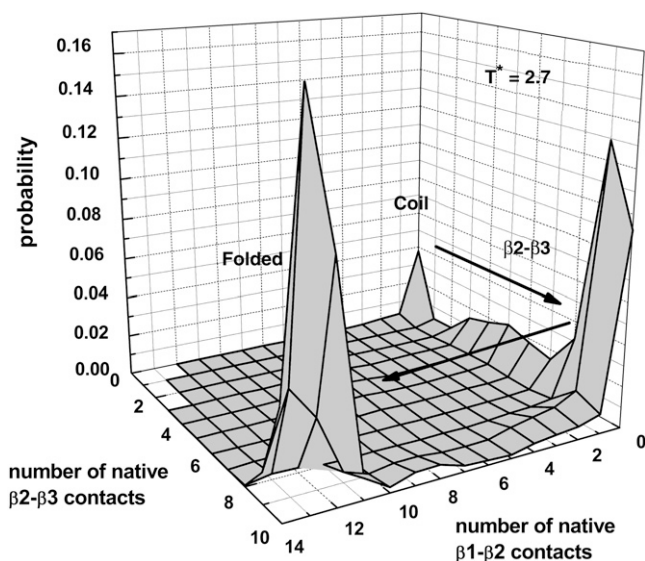


FIGURE 9 Probability distribution of the numbers of native $\beta 1$ - $\beta 2$ and $\beta 2$ - $\beta 3$ contacts at $T^* = 2.7$. Seventy-seven trajectories were run, and 2000 configurations for each trajectory were collected for statistics.

trajectories folded via the formation of hairpin $\beta 1$ - $\beta 2$. Thus, the folding pathway of the Pin1 WW domain is altered as temperature changes.

DISCUSSION

In this article, an all-atom G \ddot{o} model of Pin1 WW domain was simulated with the discontinuous MD technique. The model exhibited a two-state folding thermodynamics. The folding also exhibited a two-state folding kinetics at the transition temperature with a transition state of a mostly formed hairpin $\beta 1$ - $\beta 2$. At a low temperature, a drastically different folding behavior was, however, observed. There was a kinetic intermediate with a mostly formed hairpin 2 made of $\beta 2$ and $\beta 3$ strands. In between, two kinetic intermediates with either hairpin 1 or hairpin 2 were observed.

The drastically different folding kinetics and pathways (Fig. 10) can be understood as follows. The most entropically favorable way for forming the hairpin 1 (or the hairpin 2) is the formation of loop 1 (or loop 2) that brings $\beta 1$ and $\beta 2$ (or $\beta 2$ and $\beta 3$) together. Local contacts of loop 2 are more likely to form than those of loop 1 because loop 2 is shorter (32). However, the formation of the hairpin 2 is unlikely to complete folding by recruiting $\beta 1$, because at high temperature, their intradomain interactions are not strong enough to survive for a relative long time needed for the formation of the long loop 1. Moreover, loop 2 requires long-range interactions to be stable (Fig. 1 C). Thus, the transition state is made of more stable hairpin $\beta 1$ - $\beta 2$ instead of hairpin $\beta 2$ - $\beta 3$, although longer time is required for the formation of local contacts in loop 1. Once the hairpin 1 forms, it is stabilized

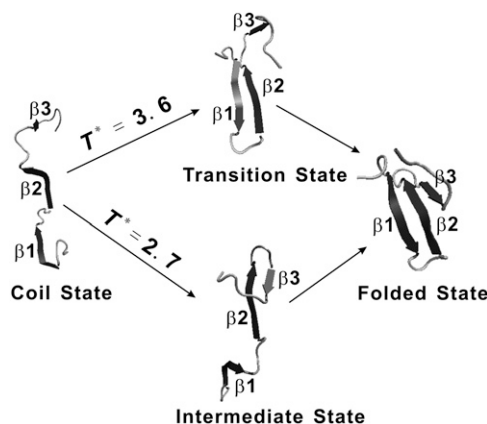


FIGURE 10 A schematic presentation of the two folding pathways at approximately the transition temperature and at a low temperature.

by the hydrophobic contacts and H-bonds between strands $\beta 1$ and $\beta 2$, and it is relatively easy to bring $\beta 3$ to fold because of a short loop 2. Situation is, however, different at low temperatures. The hairpin $\beta 2$ - $\beta 3$ is, once formed, stable against thermal fluctuation. It becomes a kinetic intermediate while waiting for the formation of the long loop 1 to bring $\beta 1$ to fold (Figs. 9 and 10), although thermodynamically, this protein is a two-state protein (Figs. 2 and 3). Above explanation reflects that folding results from an interplay between entropy (local contacts form first) and enthalpy (interaction). At approximately that transition temperature, interaction energy guides the folding because only the most stable region can survive thermal fluctuation to yield a productive folding event. At a low temperature, some early local contacts produce a kinetic trap.

Recent studies devote significant efforts to find the relationship between the folding rates and various physical parameters of the structures of native proteins (55–59). These studies suggest that topology and the number of folded residues rather than detailed energetic interactions play dominant roles in determining the folding rates of small proteins (60–63). Bai et al. have found that the folding rate is correlated to the critical nucleation size of the transition state more than that of the native structure (64). Here, we find that the TSE of the Pin1 WW domain has a critical nucleation size of 18 if the criterion of $\phi_i^{\text{sim}} > 0.4$ is used to define the critical nucleation residues. This size from our simulation is in a reasonable agreement with that from the previous studies, i.e., 16, based on a simple analytical model (64). Our simulations reveal that the nuclear residues include residues 9–20 and residues 25–29. These residues are mainly located at loop 1, $\beta 1$ and $\beta 2$.

The above described simulation results are consistent with several experimental observations. It is known that the Pin1 WW domain folds in a two-state behavior thermodynamically and kinetically (26). At approximately temperature 45°C, which is slightly lower than the transition temperature

(59°C), the ϕ^{exp} values in strands $\beta 1$ and $\beta 2$ are higher than ϕ^{exp} in $\beta 3$ strand. This suggests that the hairpin 1 is more nativelike than the hairpin 2 in the TSE (26). The A-to-E mutation also shows that the transition state has a mostly formed backbone structure for hairpin 1 but a yet flexible $\beta 3$ strand (27). Our simulation also predicted the different pathways of the Pin1 WW domain. The temperature dependence, however, is yet to be tested experimentally. Experimental temperature dependence so far was conducted at temperature near or higher than transition temperature (26,27). This region, however, is very difficult to perform simulations in, because folding rarely occurs at temperatures higher than transition temperatures.

To the best of our knowledge, this work is the first detailed folding kinetic study of Pin1 WW domain on the all-atom level. Using C_{α} models, Karanicolas et al. (19) predicted that the hairpin 2 of Pin1 WW domain folds before hairpin 1 while Cecconi et al. (20) suggested a zipping mechanism after the formation of a collapsed hydrophobic core. However, our study shows that the hairpin 1 is formed before the hairpin 2 near the transition temperature. Our result is consistent with experimental observations by Gruebele et al. (26,27). Previously, we have argued that it is difficult to reproduce the specific folding behavior of a β -hairpin by a C_{α} model, while it is relatively easy by using an all-atom model (34). This study confirms the above finding, that is, packing and native atomic contact interactions play an important role in the folding mechanism of proteins. (Certainly, it is possible that a sophisticated C_{α} -based model may be parameterized to reproduce the behavior of an all-atom model.) A similar conclusion was obtained in a separate folding kinetic study of a three-helix bundle fragment B of Staphylococcal protein A (33).

Interestingly, some theoretical studies of other WW domains agree with our findings on Pin1 WW domain. For example, the simulations of YAP 65 (15,16) and FBP 28 (22) indicate that the $\beta 2$ and $\beta 3$ strands of hairpin 2 are dissociated first in unfolding. This can be considered as the reverse of the folding process of Pin1 WW domain observed here. The transition state of FBP 28 also has a well-established hairpin 1 and flexible strand 3 (21,22). Moreover, loop 1 is more flexible compared to the contacts between $\beta 1$ and $\beta 2$ strands (21). Some misfolded intermediate states exist at low temperatures (22). These results are consistent with our results on Pin1 WW domain.

The authors are grateful for financial support from the Natural Science Foundation of China (grants No. 50533010 and No. 20574013, and a Two-Base grant); the Chinese Ministry of Education (Key grant No. 305004); the 973 Project (grant No. 2005CB522700); Science and Technology Developing Foundation of Shanghai (grant No. 055207082); and the Senior Visiting Scholar Foundation of the Key Laboratory. Computer support from National High Performance Computing Center of Shanghai is acknowledged. Y.Z. also thanks the National Institutes of Health, Bethesda, MD (grants No. R01 GM 966049 and No. R01 GM 068530), and the Howard Hughes Medical Institute, SUNY Buffalo and the Center for Computational Research and the Keck Center for Computational Biology, SUNY Buffalo for their grants.

REFERENCES

- Ferrara, P., and A. Caflisch. 2000. Folding simulations of a three-stranded antiparallel β -sheet peptide. *Proc. Natl. Acad. Sci. USA.* 97:10780–10785.
- Pande, V. S., and D. S. Rokhsar. 1999. Molecular dynamics simulations of unfolding and refolding of a β -hairpin fragment of protein G. *Proc. Natl. Acad. Sci. USA.* 96:9062–9067.
- Soto, P., and G. Colombo. 2004. Characterization of the conformational space of a triple-stranded β -sheet forming peptide with molecular dynamics simulations. *Proteins.* 57:734–746.
- García, A. E., and K. Y. Sanbonmatsu. 2001. Exploring the energy landscape of a β -hairpin in explicit solvent. *Proteins.* 42:345–354.
- Seibert, M. M., A. Patriksson, B. Hess, and D. van der Spoel. 2005. Reproducible polypeptide folding and structure prediction using molecular dynamics simulations. *J. Mol. Biol.* 354:173–183.
- Lwin, T. Z., and R. Luo. 2006. Force field influences in β -hairpin folding simulations. *Protein Sci.* 15:2642–2655.
- Guo, W. H., S. Lampoudi, and J. E. Shea. 2003. Posttransition state desolvation of the hydrophobic core of the src-SH3 protein domain. *Biophys. J.* 85:61–69.
- Marianayagam, N. J., and S. E. Jackson. 2004. The folding pathway of ubiquitin from all-atom molecular dynamics simulations. *Biophys. Chem.* 111:159–171.
- Cheung, M. S., A. E. García, and J. N. Onuchic. 2002. Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc. Natl. Acad. Sci. USA.* 99:685–690.
- Gosavi, S., L. L. Chavez, P. A. Jennings, and J. N. Onuchic. 2006. Topological frustration and the folding of interleukin- β . *J. Mol. Biol.* 357:986–996.
- Sudol, M., K. Sliwa, and T. Russo. 2001. Functions of WW domains in the nucleus. *FEBS Lett.* 490:190–195.
- Sudol, M., and T. Hunter. 2000. New wrinkles for an old domain. *Cell.* 103:1001–1004.
- Zarrinpar, A., and W. A. Lim. 2000. Converging on proline: the mechanism of WW domain peptide recognition. *Nat. Struct. Biol.* 7:611–613.
- Macias, M. J., M. Hyvonen, E. Baraldi, J. Schultz, M. Sudol, M. Saraste, and H. Oschkinat. 1996. Structure of the WW domain of a kinase-associated protein complexed with a proline-rich peptide. *Nature.* 382:646–649.
- Ibragimova, G. T., and R. C. Wade. 1999. Stability of the β -sheet of the WW domain: a molecular dynamics simulation study. *Biophys. J.* 77:2191–2198.
- Ferguson, N., J. R. Pires, F. Toepert, C. M. Johnson, Y. P. Pan, R. Volkmer-Engert, J. Schneider-Mergener, V. Daggett, H. Oschkinat, and A. Fersht. 2001. Using flexible loop mimetics to extend ϕ -value analysis to secondary structure interactions. *Proc. Natl. Acad. Sci. USA.* 98:13008–13013.
- Karanicolas, J., and C. L. Brooks. 2003. The structural basis for biphasic kinetics in the folding of the WW domain from a formin-binding protein: lessons for protein design? *Proc. Natl. Acad. Sci. USA.* 100:3954–3959.
- Karanicolas, J., and C. L. Brooks. 2004. Integrating folding kinetics and protein function: biphasic kinetics and dual binding specificity in a WW domain. *Proc. Natl. Acad. Sci. USA.* 101:3432–3437.
- Karanicolas, J., and C. L. Brooks. 2003. Improved Gō-like models demonstrate the robustness of protein folding mechanisms towards non-native interactions. *J. Mol. Biol.* 334:309–325.
- Cecconi, F., C. Guardiani, and R. Livi. 2006. Testing simplified proteins models of the hPin1 WW domain. *Biophys. J.* 91:694–704.
- Petrovich, M., A. L. Jonsson, N. Ferguson, V. Daggett, and A. R. Fersht. 2006. ϕ -Analysis at the experimental limits: mechanism of β -hairpin formation. *J. Mol. Biol.* 360:865–881.
- Mu, Y. G., L. Nordenskiöld, and J. P. Tam. 2006. Folding, misfolding, and amyloid protofibril formation of WW domain FBP28. *Biophys. J.* 90:3983–3992.

23. Ferguson, N., C. M. Johnson, M. Macias, H. Oschkinat, and A. Fersht. 2001. Ultrafast folding of WW domains without structured aromatic clusters in the denatured state. *Proc. Natl. Acad. Sci. USA*. 98:13002–13007.
24. Kaul, R., A. R. Angeles, M. Jäger, E. T. Powers, and J. W. Kelly. 2001. Incorporating β -turns and a turn mimetic out of context in loop 1 of the WW domain affords cooperatively folded β -sheets. *J. Am. Chem. Soc.* 123:5206–5212.
25. Macias, M. J., V. Gervais, C. Civera, and H. Oschkinat. 2000. Structural analysis of WW domains and design of a WW prototype. *Nat. Struct. Biol.* 7:375–379.
26. Jäger, M., H. Nguyen, J. C. Crane, J. W. Kelly, and M. Gruebele. 2001. The folding mechanism of a β -sheet: the WW domain. *J. Mol. Biol.* 311:373–393.
27. Deechongkit, S., H. Nguyen, E. T. Powers, P. E. Dawson, M. Gruebele, and J. W. Kelly. 2004. Context-dependent contributions of backbone hydrogen bonding to β -sheet folding energetics. *Nature*. 430:101–105.
28. Ranganathan, R., K. P. Lu, T. Hunter, and J. P. Noel. 1997. Structural and functional analysis of the mitotic rotamase Pin1 suggests substrate recognition is phosphorylation dependent. *Cell*. 89:875–886.
29. Kowalski, J. A., K. Liu, and J. W. Kelly. 2002. NMR solution structure of the isolated Apo Pin1 WW domain: comparison to the x-ray crystal structures of Pin1. *Biopolymers*. 63:111–121.
30. Brown, S., N. J. Fawzi, and T. Head-Gordon. 2003. Coarse-grained sequences for protein folding and design. *Proc. Natl. Acad. Sci. USA*. 100:10712–10717.
31. Taketomi, H., Y. Ueda, and N. Gō. 1975. Studies on protein folding, unfolding and fluctuations by computer-simulation. I. Effect of specific amino-acid sequence represented by specific inter-unit interactions. *Int. J. Pept. Protein Res.* 7:445–459.
32. Linhananta, A., H. Y. Zhou, and Y. Q. Zhou. 2002. The dual role of a loop with low loop contact distance in folding and domain swapping. *Protein Sci.* 11:1695–1701.
33. Linhananta, A., and Y. Q. Zhou. 2002. The role of side chain packing and native contact interactions in folding: discontinuous molecular dynamics folding simulations of an all-atom Gō model of fragment B of Staphylococcal protein A. *J. Chem. Phys.* 117:8983–8995.
34. Zhou, Y. Q., and A. Linhananta. 2002. Role of hydrophilic and hydrophobic contacts in folding of the second β -hairpin fragment of protein G: molecular dynamics simulation studies of an all-atom model. *Proteins*. 47:154–162.
35. Zhou, Y. Q., and A. Linhananta. 2002. Thermodynamics of an all-atom off-lattice model of the fragment B of Staphylococcal protein A: implication for the origin of the cooperativity of protein folding. *J. Phys. Chem. B*. 106:1481–1485.
36. Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM—a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 4:187–217.
37. Neria, E., S. Fischer, and M. Karplus. 1996. Simulation of activation free energies in molecular systems. *J. Chem. Phys.* 105:1902–1921.
38. Weeks, J. D., D. Chandler, and H. C. Andersen. 1971. Role of repulsive forces in determining the equilibrium structure of simple liquids. *J. Chem. Phys.* 54:5237–5247.
39. Sherwood, A. E., and J. M. Prausnitz. 1964. Intermolecular potential functions and the second and third virial coefficients. *J. Chem. Phys.* 41:429–437.
40. Andersen, H. C. 1980. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.* 72:2384–2393.
41. Zhou, Y. Q., M. Karplus, J. M. Wichert, and C. K. Hall. 1997. Equilibrium thermodynamics of homopolymers and clusters: molecular dynamics and Monte Carlo simulations of systems with square-well interactions. *J. Chem. Phys.* 107:10691–10708.
42. Zhou, Y. Q., and M. Karplus. 1999. Folding of a model three-helix bundle protein: a thermodynamic and kinetic analysis. *J. Mol. Biol.* 293:917–951.
43. Ferrenberg, A. M., and R. H. Swendsen. 1989. Optimized Monte Carlo data analysis. *Phys. Rev. Lett.* 63:1195–1198.
44. Fersht, A. R., A. Matouschek, and L. Serrano. 1992. The folding of an enzyme. 1. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* 224:771–782.
45. Paci, E., M. Vendruscolo, C. M. Dobson, and M. Karplus. 2002. Determination of a transition state at atomic resolution from protein engineering data. *J. Mol. Biol.* 324:151–163.
46. Paci, E., C. T. Friel, K. Lindorff-Larsen, S. E. Radford, M. Karplus, and M. Vendruscolo. 2004. Comparison of the transition state ensembles for folding of Im7 and Im9 determined using all-atom molecular dynamics simulations with ϕ -value restraints. *Proteins*. 54:513–525.
47. Geierhaas, C. D., R. B. Best, E. Paci, M. Vendruscolo, and J. Clarke. 2006. Structural comparison of the two alternative transition states for folding of TI I27. *Biophys. J.* 91:263–275.
48. Hammond, G. S. 1955. A correlation of reaction rates. *J. Am. Chem. Soc.* 77:334–338.
49. Sánchez, I. E., and T. Kiefhaber. 2003. Hammond behavior versus ground state effects in protein folding: evidence for narrow free energy barriers and residual structure in unfolded states. *J. Mol. Biol.* 327:867–884.
50. Ding, F., N. V. Dokholyan, S. V. Buldyrev, H. E. Stanley, and E. I. Shakhnovich. 2002. Direct molecular dynamics observation of protein folding transition state ensemble. *Biophys. J.* 83:3525–3532.
51. Clementi, C., H. Nymeyer, and J. N. Onuchic. 2000. Topological and energetic factors: what determines the structural details of the transition state ensemble and “en-route” intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* 298:937–953.
52. Cho, S. S., Y. Levy, and P. G. Wolynes. 2006. P versus Q: structural reaction coordinates capture protein folding on smooth landscapes. *Proc. Natl. Acad. Sci. USA*. 103:586–591.
53. Wang, J., K. Zhang, H. Y. Lu, and E. K. Wang. 2005. Quantifying kinetic paths of protein folding. *Biophys. J.* 89:1612–1620.
54. Du, R., V. S. Pande, A. Y. Grosberg, T. Tanaka, and E. S. Shakhnovich. 1998. On the transition coordinate for protein folding. *J. Chem. Phys.* 108:334–350.
55. Plaxco, K. W., K. T. Simons, and D. Baker. 1998. Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* 277:985–994.
56. Kuznetsov, I. B., and S. Rackovsky. 2004. Class-specific correlations between protein folding rate, structure-derived, and sequence-derived descriptors. *Proteins*. 54:333–341.
57. Pintar, A., and S. Pongor. 2005. The “first in-last out” hypothesis on protein folding revisited. *Proteins*. 60:584–590.
58. Gromiha, M. M., and S. Selvaraj. 2001. Comparison between long-range interactions and contact order in determining the folding rate of two-state proteins: application of long-range order to folding rate prediction. *J. Mol. Biol.* 310:27–32.
59. Zhou, H. Y., and Y. Q. Zhou. 2002. Folding rate prediction using total contact distance. *Biophys. J.* 82:458–463.
60. Baker, D. 2000. A surprising simplicity to protein folding. *Nature*. 405:39–42.
61. Plaxco, K. W., K. T. Simons, I. Ruczinski, and D. Baker. 2000. Topology, stability, sequence, and length: defining the determinants of two-state protein folding kinetics. *Biochemistry*. 39:11177–11183.
62. Makarov, D. E., C. A. Keller, K. W. Plaxco, and H. Metiu. 2002. How the folding rate constant of simple, single-domain proteins depends on the number of native contacts. *Proc. Natl. Acad. Sci. USA*. 99:3535–3539.
63. Paci, E., K. Lindorff-Larsen, C. M. Dobson, M. Karplus, and M. Vendruscolo. 2005. Transition state contact orders correlate with protein folding rates. *J. Mol. Biol.* 352:495–500.
64. Bai, Y. W., H. Y. Zhou, and Y. Q. Zhou. 2004. Critical nucleation size in the folding of small apparently two-state proteins. *Protein Sci.* 13: 1173–1181.
65. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD—visual molecular dynamics. *J. Mol. Graph.* 14:33–38.